

# Data-adaptive unfolding of non-standard random-matrix spectra: cross-over of long-range correlation statistics

R. Fossion<sup>\*,1,2</sup> G. Torres Vargas,<sup>3</sup> S. Díaz Gómez,<sup>3</sup> and J.C. López Vieyra<sup>4</sup>

<sup>1</sup>*Instituto Nacional de Geriatria, Periférico Sur No. 2767, 10200 México D.F., Mexico*

<sup>2</sup>*Centro de Ciencias de la Complejidad (C3), Universidad Nacional Autónoma de México, 04510 México D.F., Mexico*

<sup>3</sup>*Posgrado en Ciencias Físicas, Universidad Nacional Autónoma de México, 04510 México D.F., Mexico*

<sup>4</sup>*Instituto de Ciencias Nucleares, Universidad Nacional Autónoma de México, 04510 México D.F., Mexico*

Recently, a model-free and data-adaptive unfolding method, based on normal-mode analysis, was proposed and demonstrated for standard Gaussian ensembles of Random Matrix Theory (RMT). Here, the method is applied to a sparse matrix ensemble and to the  $\nu$ -Hermite ensemble. In correspondence with known results, level fluctuations are obtained that exhibit a cross-over between soft and rigid behaviour, whereas fluctuations of standard RMT ensembles are scale invariant. Possible artefacts introduced by standard unfolding techniques are avoided. Ensemble-averaged and individual-spectrum averaged statistics are calculated consistently within the same basis of normal modes and allow to characterize the ergodicity of the ensemble under study.

PACS numbers: 05.45.Tp, 05.45.Mt, 89.75.-k, 02.50.Sk, 02.10.Yn

## I. INTRODUCTION

Standard Gaussian ensembles from Random Matrix Theory (RMT) [1] have been enormously successful in the modelling of the fluctuations of quantum excitation spectra [2], and recently have been used as well in multivariate statistics to model the fluctuations of eigenspectra of correlation matrices of complex systems [3]. More general random-matrix ensembles introduce new statistical features that are absent in standard RMT, such as Gaussian instead of semicircular global eigenvalue densities [4], breakdown of the scale invariance of the long-range fluctuation statistics [5–7] and nonergodicity [4, 5, 8]. One of the difficulties in the statistical study of spectral fluctuations is the *unfolding* procedure which serves to separate the global density of eigenlevels  $\bar{\rho}(E)$  from the local fluctuations  $\tilde{\rho}(E) = \rho(E) - \bar{\rho}(E)$ . The unfolding procedure is not trivial, and statistical results are sensitive to the unfolding applied [9, 10].

In a previous publication [11], we proposed an unfolding method that is data-adaptive and model-free, and that expresses a spectrum in an exact way as the superposition of global and fluctuation normal modes. When the unfolding method is applied to standard Gaussian ensembles, the normal modes associated to the fluctuations are scale invariant and obey specific power laws that distinguish between soft and rigid spectra. In the present contribution, we apply the data-adaptive unfolding technique to a sparse matrix model and to the  $\nu$ -Hermite ensemble. Scale invariance is lost for the fluctuations, with a cross-over between soft and rigid properties at different scales, in correspondence with known results [5–7], but without possible artifacts introduced by standard unfolding techniques.

## II. DATA-ADAPTIVE UNFOLDING

Following the discussion of Ref. [11], we consider an ensemble of  $m = 1 \dots M$  eigenspectra  $E^{(m)}(n)$ , where each spectrum consists of  $n = 1 \dots N$  levels. Each spectrum is conveniently accommodated in a row of the  $M \times N$  dimensional matrix  $\mathbf{X}$ , which can now be interpreted as a multivariate time series,

$$\mathbf{X} = \begin{pmatrix} E^{(1)}(1) & E^{(1)}(2) & \dots & E^{(1)}(N) \\ E^{(2)}(1) & E^{(2)}(2) & \dots & E^{(2)}(N) \\ \vdots & \vdots & \ddots & \vdots \\ E^{(M)}(1) & E^{(M)}(2) & \dots & E^{(M)}(N) \end{pmatrix}. \quad (1)$$

Singular Value Decomposition (SVD) decomposes  $\mathbf{X}$  in a *unique* and *exact* way as,

$$\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T = \sum_{k=1}^r \sigma_k \mathbf{X}_k = \sum_{k=1}^r \sigma_k \vec{u}_k \vec{v}_k^T, \quad (2)$$

where  $\mathbf{\Sigma}$  is an  $M \times N$ -dimensional matrix with only diagonal elements which are ordered *singular values* or weights  $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r$ , with  $r \leq \text{Min}[M, N] = \text{rank}(\mathbf{X})$ . The vectors  $\vec{v}_k$  are orthonormal, and correspond to the  $k$ th columns of the  $N \times N$ -dimensional matrix  $\mathbf{V}$ . They are *normal modes*, and can be interpreted as data-adaptive filters corresponding to different scales, that are produced by the data themselves. They constitute a basis for the row space of  $\mathbf{X}$ . The vectors  $\vec{u}_k$  are orthonormal as well, corresponding to the  $k$ th columns of the  $M \times M$ -dimensional matrix  $\mathbf{U}$ , and can be interpreted as the associated *projection coefficients*. The elementary matrices  $\mathbf{X}_k = \vec{u}_k \vec{v}_k^T \equiv \vec{u}_k \otimes \vec{v}_k$  are calculated from the outer product of  $\vec{u}_k$  and  $\vec{v}_k$ . In this way, any matrix row of  $\mathbf{X}$  containing a particular excitation spectrum  $E^{(m)}(n)$  can be written as,

$$E^{(m)}(n) = \sum_{k=1}^r \sigma_k U_{mk} \vec{v}_k^T(n). \quad (3)$$

\*fossion@nucleares.unam.mx

The square of a singular value or weight gives a partial variance,  $\lambda_k = \sigma_k^2$ , which is the fraction of the total variance  $\lambda_{\text{tot}} = \sum_{k=1}^r \lambda_k$  of the multivariate time series  $\mathbf{X}$  carried by the normal mode  $\vec{v}_k$ . A first result of Ref. [11] is that the main feature of a typical excitation spectrum or matrix eigenspectrum  $E(n)$  is a dominant ascending trend  $\bar{E}(n)$ , whereas the superposed fluctuations  $\tilde{E}(n) = E(n) - \bar{E}(n)$  are usually orders of magnitude smaller. Consequently, the global properties of a spectrum are represented by non-oscillating normal modes  $\vec{v}_k$  associated to the first few  $k = 1, \dots, n_T$  dominant partial variances  $\lambda_k$  in the so-called *scree diagram* of ordered partial variances, which can be distinguished easily from the much smaller higher-order partial variances  $\lambda_k$  with  $k = n_T + 1, \dots, r$  corresponding to oscillating normal modes  $\vec{v}_k$  associated to the fluctuations.

Another result from Ref. [11] is that when the data-adaptive unfolding of Eq. (2) is applied to the standard Gaussian ensembles of RMT, the part of the scree diagram associated to the fluctuation normal modes obeys the power law,

$$\lambda_k \propto 1/k^\gamma, \quad (4)$$

indicating that the long-range correlations of the fluctuations are scale invariant (fractal). The exponent  $\gamma$  characterizes the rigidity of the fluctuations, with  $\gamma = 2$  in the case of soft spectra in the Poisson limit, and  $\gamma = 1$  in the case of rigid spectra in the limit of the Gaussian Orthogonal Ensemble (GOE). On the other hand, any particular spectrum from the ensemble can be unfolded separately using Eq. (3), after which its individual fluctuations can be studied with, e.g., Fourier spectral analysis. In the case of the standard Gaussian ensembles, the Fourier power spectrum of the fluctuations of the individual eigenspectra also obeys a power law,

$$P(f) \propto 1/f^\beta, \quad (5)$$

with  $\beta = \gamma = 2$  (Poisson) and  $\beta = \gamma = 1$  (GOE), such that the scale invariance property does not seem to depend on the particular basis used for the decomposition. In the following, we will apply the data-adaptive unfolding to the cases of a sparse-matrix ensemble and the  $\nu$ -Hermite ensemble.

### III. APPLICATION TO A SPARSE RANDOM-MATRIX ENSEMBLE

In Ref. [5], a new ensemble of sparse real symmetric matrices was proposed starting from GOE, using a sparsity parameter  $s$  which is the fraction of the  $N(N-1)/2$  independent off-diagonal matrix elements chosen to be nonvanishing. All diagonal elements are kept nonzero. The nonvanishing matrix elements  $H_{ij}$  are chosen independently and at random from a Gaussian distribution

$\mathcal{N}(0, \sqrt{2}\sigma)$  with mean 0 and standard deviation  $\sqrt{2}\sigma$  for the diagonal elements, and  $\mathcal{N}(0, \sigma)$  for the off-diagonal ones,

$$P(H_{ij}) = \frac{1}{\sqrt{2\pi\sigma_{ij}}} \exp\left(-\frac{H_{ij}^2}{2\sigma_{ij}^2}\right), \quad (6)$$

with  $\sigma_{ij} = 1 + \delta_{ij}$ . GOE statistics is recovered in the limiting case of maximum sparsity  $s = 1$  and Poisson statistics in the case of minimum sparsity  $s = 0$ . For arbitrary sparsity, the spectral density  $\rho(E)$  is intermediate between a Gaussian shape and a semicircle, and is difficult to describe analytically. Thus, Ref. [5] carried out a numerical unfolding, fitting a polynomial of arbitrary degree to the integrated level density of either a single realization of the eigenspectrum (individual-spectrum unfolding or self unfolding), or to the integrated level density averaged over the whole ensemble (ensemble unfolding). In the case of *ergodic* ensembles, both types of unfolding give the same statistical results, which is however not the case for *nonergodic* ensembles. After this prior unfolding, a normal-mode analysis similar to Eqs. (1)-(2) was applied to the ensemble of fluctuations, in which case only fluctuation normal modes and no trend normal modes are obtained. For intermediate sparsities  $0 < s < 1$ , instead of the scale invariance property of Eq. (4), a cross-over was observed for the fluctuations with rigid behaviour ( $\gamma = 1$ ) for higher order-numbers  $k$  (finer scales) and soft behaviour ( $\gamma = 2$ ) at lower order-numbers (coarser scales). The location of the cross-over  $k_\times$  was found to be proportional with the dimension of the spectrum,  $k_\times \propto \sqrt{N}$ , and shifts to higher order-numbers  $k$  (finer scales) for lower sparsities  $s$ . An ambiguity was found in the number variance fluctuation measure  $\Sigma^2$  when calculated after individual-spectrum unfolding or after ensemble unfolding, which however was argued to be an artificial effect of the two types of unfolding that were compared, and Ref [5] warned against interpreting erroneously the sparse matrix model of being nonergodic.

In Fig. 1 (panel a), we show results for the data-adaptive unfolding of Eq. (2) applied to an ensemble of  $M = 500$  spectra of sparse matrices, each spectrum containing  $N = 2000$  levels. The optimal ensemble size is  $M \approx N/4$ , because for  $M \gg N/4$  there is a long tail of insignificant partial variances in the scree diagram, whereas for  $M \ll N/4$  the range of scales of the scree diagram becomes restrained, but the statistical properties do not depend on the particular choices of  $N$  and  $M$  [11]. The scree diagram clearly distinguishes between the first  $k = 1, 2$  trend modes and  $k = 3, \dots, r$  higher-order fluctuations modes, with  $r = 500$ . The fluctuation part of the scree diagram exhibits the cross-over behaviour, and  $k_\times$  shifts towards higher order-numbers (finer scales) for decreasing sparsity, in correspondence to Ref. [5]. On the other hand, after individual unfolding of the separate spectra of the ensemble, according to

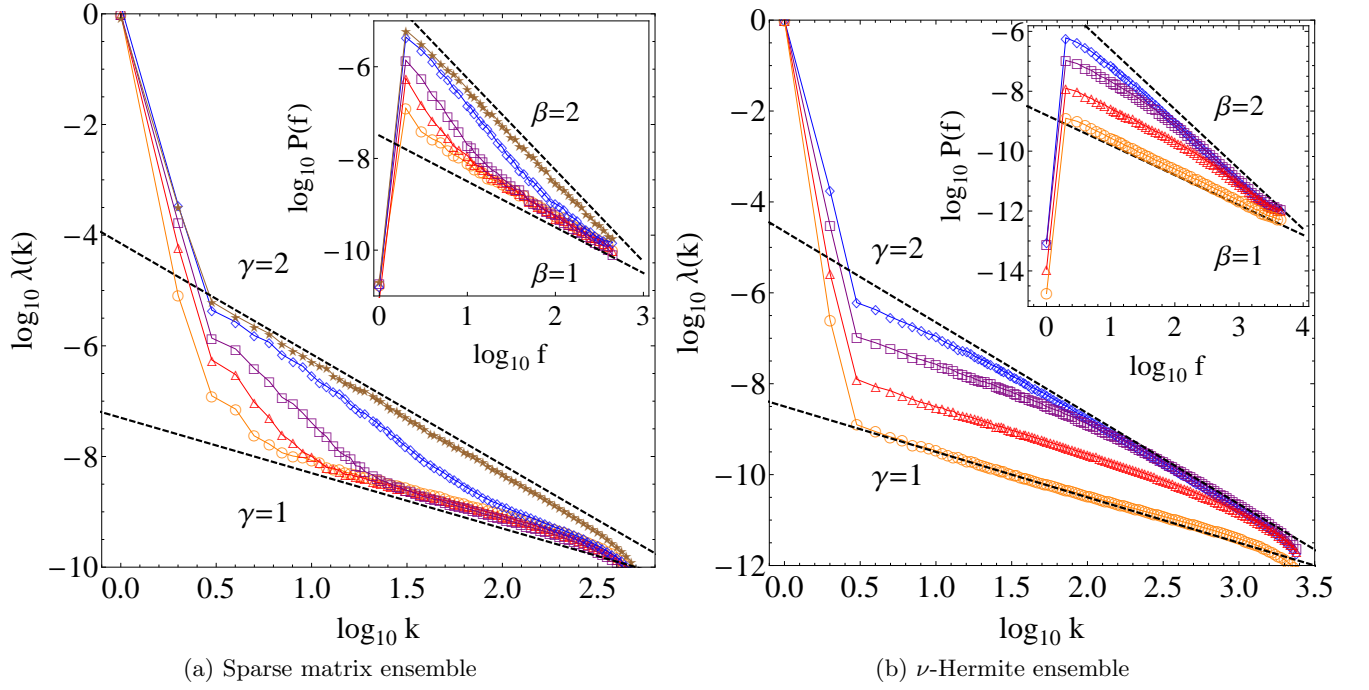


FIG. 1: (a) Scree diagram of ordered partial variances  $\lambda_k$  of an ensemble of  $M = 500$  sparse matrices of dimension  $N = 2000$ , see Sect. III, for sparsities  $s = 0.025$  (circles),  $s = 0.01$  (triangles),  $s = 0.005$  (squares),  $0.0025$  (diamonds) and  $s = 0.0005$  (asterisks). Using the same symbols, the inset displays the Fourier power spectrum  $P(f)$  of the fluctuations of the separate eigenspectra, averaged over the whole ensemble after individual unfolding is applied. (b) Similar to panel (a), but for a  $\nu$ -Hermite ensemble of  $M = 2500$  spectra with  $N = 10^4$  levels each, see Sect. IV, for  $\nu = 1.0$  (circles),  $\nu = 0.1$  (triangles),  $\nu = 0.01$  (squares) and  $\nu = 0.001$  (diamonds). In both panels, the total variance of the ensemble has been rescaled to unit variance  $\sum_k \lambda_k = 1$  to allow comparison between ensembles with different  $s$  or  $\nu$ . (Colour online).

Eq. (3), the same cross-over behaviour is observed in the Fourier power spectrum (note that the lowest frequency is the constant DC term and does not belong to the fluctuations). The scree diagram is an ensemble-averaged property and shows the collective behaviour of the normal modes common to the whole ensemble, whereas the Fourier power spectrum of the individual fluctuations is an individual-spectrum averaged property. Thus, in the present framework of data-adaptive unfolding, ensemble-averaged and individual-spectrum averaged statistics can be studied consistently within the same basis of normal modes, and avoid the ambiguities of the standard unfolding applied in Ref. [5]. Both predictions are found to correspond, and confirm the ergodicity of the sparse matrix ensemble studied. The study of nonergodic ensembles within the present framework will be the topic of another contribution [12].

#### IV. APPLICATION TO THE $\nu$ -HERMITE RANDOM-MATRIX ENSEMBLE

The  $\nu$ -Hermite ensemble or  $\nu$ -Gaussian ensemble is also called the continuous Gaussian ensemble because the parameter  $\nu$  interpolates continuously between the classical Gaussian ensembles of RMT, e.g., between the Poisson limit ( $\nu = 0$ ) and GOE ( $\nu = 1$ ) [13]. One of the most convenient characteristics of the  $\nu$ -Hermite ensemble is its simple tridiagonal form, which has the important advantage of an unrivaled speedup and efficiency in numerical simulations with large matrix dimensions [6, 7]. A tridiagonal  $N \times N$  random matrix from the  $\nu$ -Hermite ensemble is real and symmetric, and can be defined as,

$$\mathbf{A}_{N,\nu} = \sigma \mathbf{H}_{N,\nu} = \sigma \begin{pmatrix} H_{11} & H_{12}/\sqrt{2} & 0 & \dots & 0 \\ H_{12}/\sqrt{2} & H_{22} & H_{23}/\sqrt{2} & 0 & \dots \\ 0 & H_{32}/\sqrt{2} & \dots & \dots & 0 \\ \dots & 0 & \dots & H_{N-1,N-1} & H_{N-1,N}/\sqrt{2} \\ 0 & \dots & 0 & H_{N,N-1}/\sqrt{2} & H_{N,N} \end{pmatrix}, \quad (7)$$

where  $\sigma$  is a scale factor, and which is chosen as  $\sigma = 1$  here. The  $2N - 1$  distinct matrix elements are independent random variables. The  $N$  diagonal elements are independently distributed standard normal random variables  $\mathcal{N}(0,1)$ . The off-diagonal elements  $H_{m,m+1}$  ( $m = 1, \dots, N - 1$ ) have a  $\chi$  distribution with  $m\nu$  degrees of freedom whose probability density is  $q_{N,\nu}(x) = 2^{1-m\nu/2} x^{m\nu-1} \exp(-x^2/2) / \Gamma(m\nu/2)$  ( $x \geq 0$ ). For  $0 < \nu < 1$ , the level density evolves from a smooth shape intermediate between a Gaussian and a Wigner semicircle to the Wigner semicircle in the limit for large matrix sizes  $N$  [14]. For finite  $N$ , it can be difficult to unfold the eigenspectrum analytically. Therefore, in [6], the unfolding was performed numerically as a polynomial fit to the ensemble-averaged level density, whereas in [7] a double unfolding was performed, first unfolding with Wigner's semicircle law, and afterwards re-unfolding by means of a fit with Chebyshev polynomials. After this prior unfolding, in [7], Fourier spectral analysis was applied to the unfolded fluctuations, and in [6], the unfolded fluctuations were studied with Daubechies wavelets and it was checked that other types of wavelets lead to the similar results. Both studies confirm that the interpolation between the Poisson and GOE limits is heterogeneous, with soft behaviour ( $1/f^2$  power spectrum) at the finest scales, and rigid behaviour ( $1/f$  power spectrum) at the coarsest scales. In the analytical calculations, the cross-over frequency is predicted to occur at  $f_\times = \nu N/2$ , whereas in the numerical calculations the cross-over is smoothed [7].

In Fig. 1 (panel b), results are shown for the data-adaptive unfolding of a  $\nu$ -Hermite ensemble of  $M = 2500$  spectra. As the cross-over becomes more obvious for large dimensions [6, 7], spectra with  $N = 10^4$  levels are chosen. Again, the scree diagram clearly distinguishes between the first  $k = 1, 2$  trend components and the  $k = 3, \dots, r$  higher-order fluctuation components,

with  $r = 2500$ . The fluctuation part of the scree diagram shows the aforementioned cross-over behaviour, which is confirmed by the Fourier power spectrum of the fluctuations of the individual eigenspectra (the lowest frequency is again the constant DC term). In the present approach, possible artifacts introduced by traditional unfolding techniques are avoided, e.g., the several points in the Fourier spectrum at low frequencies that fall far below the theoretical predictions in Ref. [7]. In the present approach, the normal modes are generated by the data themselves, avoiding the necessity to compare between different models, as in the case of studies with user-defined wavelets as in Ref. [6].

## V. CONCLUSIONS

We applied a data-adaptive unfolding to random-matrix ensembles of sparse matrices and the  $\nu$ -Hermite ensemble. The previously reported cross-over of the fluctuation statistics is obtained already during the unfolding step, and avoids possible artifacts introduced by standard unfolding techniques. Additionally, ensemble-averaged and individual-spectrum averaged statistics can be calculated consistently within the same basis of normal modes, which allows to characterize the ergodicity of the ensemble under study.

## Acknowledgements

We acknowledge financial support from CONACYT (CB-2011-01-167441 and CB-2010-01-155663) and the Instituto Nacional de Geriatria (project DI-PI-002/2012). We wish to thank the anonymous referee of our previous publication, who brought the interesting topic of ergodicity to our attention.

- 
- [1] M. L. Mehta, *Random matrices* (Acad. Press, New York, 1991), 2nd ed.
  - [2] F. Haake, *Quantum signatures of chaos* (Springer, Heidelberg, 2010), 3rd ed.
  - [3] J. Kwapień and S. Drożdż, *Phys. Rep.* **515**, 115 (2012).
  - [4] J. Flores, M. Horoi, M. Müller and T. H. Seligman, *Phys. Rev. E* **63**, 026204 (2001).
  - [5] A. D. Jackson, C. Mejia-Monasterio, T. Rupp, M. Saltzer and T. Wilke, *Nucl. Phys. A* **687**, 405 (2001).
  - [6] C. Male, G. Le Caër and R. Delannay, *Phys. Rev. E* **76**, 042101 (2007).
  - [7] A. Relaño, L. Muñoz, J. Retamosa, E. Faleiro and R. A. Molina, *Phys. Rev. E* **77**, 031103 (2008).
  - [8] T. Asaga, L. Benet, T. Rupp and H. A. Weidenmüller, *Europhys. Lett.* **56**, 340 (2001).
  - [9] J. M. G. Gómez, R. A. Molina, A. Relaño and J. Retamosa, *Phys. Rev. E* **66**, 036209 (2002).
  - [10] S. M. Abuelenin and A. Y. Abul-Magd, *Procedia CS* **12**, 69 (2012).
  - [11] R. Fossion, G. Torres Vargas and J. C. López Vieyra, *Phys. Rev. E* **88**, 060902(R) (2013).
  - [12] R. Fossion, arXiv:1401.0867 [nlin.CD].
  - [13] I. Dumitriu and A. Edelman, *J. Math. Phys.* **43**, 5830 (2002); I. Dumitriu and A. Edelman, *Ann. Inst. Henri Poincaré, Sect. A* **41**, 1083 (2005).
  - [14] G. Le Caër, C. Male and R. Delannay, *Phys. A* **383**, 190 (2007).